

Search for the standard model Higgs boson in the $WH \to \tau \nu b\bar{b}$ channel in 0.94 fb⁻¹ of $p\bar{p}$ collisions at $\sqrt{s}=1.96$ TeV

The DØ Collaboration URL: http://www-d0.fnal.gov (Dated: July 25, 2008)

A search for the standard model Higgs boson has been performed in 0.94 fb⁻¹ of $p\bar{p}$ collisions at 1.96 TeV, collected with the RunIIa DØ detector at the Fermilab Tevatron. The final state considered is a pair of b jets with a reconstructed tau and missing transverse energy as expected from the reaction $p\bar{p} \to WH \to \tau \nu b\bar{b}$. The search is also sensitive to the $ZH \to \tau \bar{\tau} b\bar{b}$ channel, when the second tau is not identified. For a Higgs boson mass of 115 GeV, we set a limit at 95% C.L. on the cross section times branching fraction of $(p\bar{p} \to H(W/Z)) \times (H \to b\bar{b})$, that is 35 times larger than the standard model value.

I. INTRODUCTION

The $p\bar{p} \to HW$ reaction, with $H \to b\bar{b}$ and $W \to \tau_h \nu_\tau$ (where τ_h indicates a hadronic decay of a tau), is one of the more difficult final states in which to search for a Higgs boson at the Fermilab Tevatron [1]. In this note a first search for this process in 0.94 fb⁻¹ is presented. A lower mass limit of 114.4 GeV was set by the LEP experiments for the SM Higgs boson from analyses of the reaction $e^+e^- \to HZ$ [2], while an upper limit of 144 GeV can be inferred from precision electroweak data [3]. Here and in the following, all limits quoted are at the 95% confidence level.

The final state topology considered in this analysis is a pair of b jets from the decay of the Higgs boson, with a hadronically decaying tau (some events with $\tau \to e\nu\nu$ are also retained in our selection) and missing transverse energy (E_T) due to the neutrinos from the W decay and tau decay products. The search is therefore also sensitive to the HZ channel, with $Z \to \tau\tau$ when the second tau from the Z decay is not detected. The main backgrounds arise from $(W/Z)b\bar{b}$, from (W/Z)+(non-b jets) due to flavor misidentification (mistagging), from top quark production, e.g., $t\bar{t} \to \tau\nu bq\bar{q}'\bar{b}$ or $t(q)\bar{b} \to \tau\nu b(q)\bar{b}$, from diboson production such as $WZ \to q\bar{q}'\tau\nu$ or $ZZ \to b\bar{b}\tau\tau$, and from multijet events produced by the strong interaction, with fake hadronic taus resulting from isolated jets with real b or mistagged light parton jets.

A kinematic selection is first applied to reject most of the multijet events. The two jets expected from the Higgs boson decay are next required to be tagged as b jets, using a neural network b-tagging algorithm [4]. Finally, discrimination between the signal and the remaining backgrounds is identified via the di-jet mass of the two leading (highest transverse momentum) jets in the event.

II. DATA AND SIMULATED SAMPLES

The DØ detector [5] consists of a silicon microstrip vertex detector and a fiber tracker, both located within a 2 T superconducting solenoidal magnet. A liquid-argon and uranium sampling calorimeter, housed in three cryostats, provides pseudorapidity $(\eta = -\ln(\tan(\frac{\theta}{2})))$, where θ is the angle relative to the beam axis) coverage out to $|\eta| = 4.2$ with projective towers of size 0.1×0.1 in (η, ϕ) . Here η is calculated with respect to the proton beam direction, and ϕ is the azimuthal angle. It is segmented longitudinally into four electromagnetic and up to five hadronic layers. Additional shower sampling is provided by scintillating tiles located at the boundaries between cryostats. Beyond the calorimeter, a muon detector consists of drift tube tracking detectors and scintillation trigger counters before and after 1.8 T iron toroids.

For this analysis, the data were recorded using a set of triggers designed to select events with jets and missing transverse energy. At the highest trigger level, the main requirement was $H_T > 30$ GeV, where $H_T = |-\Sigma \overrightarrow{p}_{Tjet}|$ is a measure of E_T based on jets only. After data quality requirements, the total integrated luminosity available for this analysis was (0.937 ± 0.057) fb⁻¹ [6].

Except for the background from multijet production, which was estimated from the data, all backgrounds from standard model (SM) processes were determined by Monte Carlo simulation. To estimate the multijet background in data the shape was taken from a sideband region of tau candidates that had passed a loose quality requirement, but failed the tight quality requirement that was used for the signal selection. These events were then scaled to fit the deficit between data and simulated events in the signal region at the preselection stage where the signal contribution was negligible and then reduced by data-determined factors for the final selection cuts.

The (W/Z)+jets processes were generated with Alpgen [7] interfaced with Pythia [8] for initial and final state radiation and for jet hadronization. A matching algorithm [9] was applied to avoid double counting in phase space regions which can be populated both by Alpgen or by Pythia. Light (u,d,s,g) and heavy (c,b) flavor production in association with W/Z were generated separately, and care was taken to avoid double counting between heavy flavor jets produced directly by Alpgen or subsequently by Pythia. For $t\bar{t}$ and for electroweak single top production, the Alpgen and comphep [10] generators interfaced with Pythia were used, respectively, while the vector-boson pair production processes were generated with Pythia. The signal processes (HW) and HZ were generated with Pythia for Higgs boson masses ranging from 105 to 145 GeV, in 10 GeV steps. In all these simulations, the CTEQ6L1 parton distribution functions were used [11].

For the (W/Z)+jets processes, the absolute normalizations were determined from the data before b-tagging, as explained below. The heavy flavor fractions were obtained using MCFM [12]. The cross sections for the other background processes were taken from Refs. [13], or calculated with MCFM, and the signal production cross sections were taken from Ref. [14].

Signal and background samples were passed through a full GEANT3-based simulation [15] of the detector geometry and response, and processed with the same reconstruction program as the data. Real events from randomly selected beam crossings were overlaid on simulated events to account for additional minimum bias interactions. The trigger conditions on jets and E_T were not included in the simulation, but trigger efficiency parameterizations were applied, as

determined from triggers based only on information from the muon detectors, therefore independent of those used in this analysis. Weight factors were further applied to compensate for residual differences between data and simulation for luminosity profile, primary vertex longitudinal distribution, electron, muon and jet identification. The jet energy calibration and resolution were adjusted in simulated events so as to match those measured in data.

III. EVENT SELECTION

The goal of the first stage selection criteria is to eliminate most of the large background from multijet events, while retaining a high efficiency for the signal. In a second stage, further enhancement of the search sensitivity is achieved by using heavy flavor tagging. The analysis was optimized for a Higgs boson mass of 115 GeV.

Hadronically decaying taus are characterized by a narrow isolated jet that is associated with three or less tracks. Three types of hadronically decaying taus are distinguished:

- Type 1: Calorimeter cluster, with one associated track and no electromagnetic (EM) subcluster. This corresponds mainly to the decay $\tau^{\pm} \to \pi^{\pm} \nu$.
- Type 2: Calorimeter cluster, with one associated track and at least one EM subcluster. This corresponds mainly to the decay $\tau^{\pm} \to \pi^{\pm} \pi^{0} \nu$.
- Type 3: Calorimeter cluster, with two or three associated tracks, with or without EM subclusters. This corresponds mainly to the decays $\tau^{\pm} \to \pi^{\pm} \pi^{\mp} (\pi^0) \nu$.

Tau decays into electrons are usually reconstructed as type-2 taus (as are electrons from non tau parents). These tau candidates are kept in the analysis, but they are separated (due to the different backgrounds) from the hadronically decaying type-2 taus using a dedicated neural network (τ_{nn_e}). Variables such as the fraction of energy deposited in the EM layers of the calorimeter are used here to provide discirimination. This neural network is trained using tau MC events as signal and electron MC events as background to produce a variable that peaks near one for hadronically decaying taus and near zero for electrons.

We require a reconstructed tau with transverse momentum as measured by the calorimeter greater than 12 GeV for type-1 taus, 10 GeV for type-2 taus and 15 GeV for type-3 taus. The transverse momentum of the track associated with the tau candidate is required to be greater than 7 GeV for type-1 taus and 5 GeV for type-2 taus. In the case of type-3 taus, the leading transverse momenta track must be greater than 7 GeV. As jets can be reconstructed as tau candidates, a set of neural networks, one for each tau type, has been developed to separate such taus from jets. These neural networks make use of input variables that exploit the tau signature such as longitudinal and transverse shower shapes and isolation in the calorimeter and the tracker. These neural networks $(\tau_{nn_{h(i=1,3)}})$, like the electron discriminating neural network, uses tau MC as signal but now with multijet data as background to produce an output that peaks near one for real taus and near zero for jets. Two distinct samples are created using this neural network:

- The tau candidate with the highest τ_{nn_h} in the event must satisfy $\tau_{nn_h} > 0.9$. "Analysis Sample".
- The tau candidate with the highest τ_{nn_h} in the event must satisfy $0.3 < \tau_{nn_h} < 0.7$. "Multijet Sample".

We require exactly two or three jets with $R(jet,\tau)>0.5$, where $R=\sqrt{(\Delta\eta)^2+(\Delta\phi)^2}$, and $p_T>20$ GeV and within $|\eta_{\rm det}|<2.5$, where $\eta_{\rm det}$ is the pseudorapidity measured from the detector center rather than from the primary interaction vertex. Furthermore, there have to be charged particle tracks associated with those jets, meeting minimal quality criteria such that the b-tagging algorithm can operate efficiently. Finally, we require missing transverse energy $(E_T)>30$ GeV.

To keep this analysis orthogonal to the search in the $HW \to b\bar{b}(e/\mu)\nu$ channel, events with high p_T leptons are removed as in Ref. [16]. We reject events with an electron that has $p_T > 15$ GeV and $|\eta_{\rm det}| < 1.1$ or $1.5 < |\eta_{\rm det}| < 2.5$. The removal of events with muons is more stringent as no sensitivity can be gained from these events. We remove events for which there is a muon with $p_T > 8$ GeV, $|\eta_{\rm det}| < 2.0$.

At this stage the events are dominated by multijet background and so kinematic and topological cuts are applied to reduce this background. The most discriminating cut was found to be a two dimensional triangle cut in $\Delta\phi(\tau, E_T)$ vs E_T . Most of the multijet events are found in the region where the tau is back to back with the missing transverse energy (likely due to events with fake E_T) as illustrated in Fig. 1. The signal sample populates the region where the tau and the missing transverse energy are close together as illustrated in Fig. 1 as the parent E_T is boosted and missing transverse energy is also produced from the hadronic tau decay products. We require E_T vertically E_T and E_T vertically E_T vertic

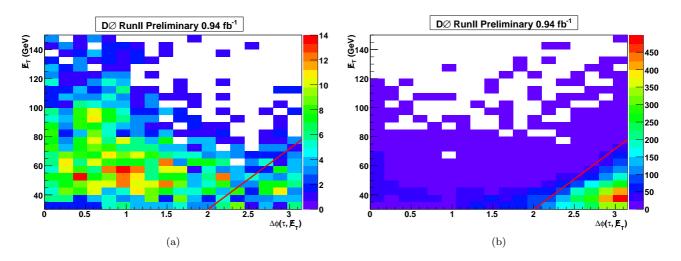


FIG. 1: Distributions for the (a) signal and (b) multijet samples of $\Delta\phi(\tau,E_T)$ vs E_T . These figures clearly show the different topology for the signal and the multijet events and how a 2D cut is used to remove the dominance of multijet events as indicated by the red line.

TABLE I: Numbers of signal and background events expected, and numbers of events observed, before and after b tagging for tau types 1 and 2 combined. The numbers for the signals are given for a Higgs boson mass of 115 GeV; "top" includes pair and single top production; VV stands for the sum of all diboson processes. The errors quoted are statistical only.

Sample	HW	HZ	W+jets	Z+jets	top	VV	$\operatorname{multijet}$	Total	Observed
before	0.534	0.042	1430	30.8	46.7	54.9	158.3	1720 ± 21	1 666
after	0.192	0.015	11.31	0.41	9.53	0.72	1.54	23.5 ± 1.05	13

The asymmetry $\mathcal{A} = (E_T - E_T)/(E_T + E_T)$ provides further discrimination between signal and multijet events. The signal is peaked around zero, however events with mismeasured E_T are smeared across all values. We require that $-0.1 < \mathcal{A} < 0.2$. In signal events, the missing track- p_T , T_T , defined as the opposite of the vectorial sum of the charged particle transverse momenta, is expected to point in a direction similar to that of the E_T , while this is not expected in multijet events where a jet energy is mismeasured. We require $\Delta \phi(E_T, T_T) < \pi/2$. The scalar sum of jet energies (H_T) is used to discriminate the signal sample from $t\bar{t}$ events that become important after b-tagging. The H_T (from good jets alone and not the tau) in the event is required to be less than 200 GeV. To remove events where the di-jet mass is far away from that of the signal we are sensitive to we require 50 GeV $< m_{jj} < 200$ GeV.

After the selection cuts we find that type-3 tau events are dominated by multijet background and offer very little improvement of the sensitivity so only type-1 and type-2 taus are considered from here onward.

Agreement between data and background is reasonable after these selection cuts (Fig. 2), so we take advantage of the large branching fraction for $H \to b\bar{b}$ by requiring that the two leading jets be b tagged. Asymmetric cuts on the outputs of the b tagging algorithm neural network [4] were applied to the two leading jets. They were chosen such that one of those jets is tagged with an efficiency of 73%, and the other with an efficiency of 48%, for jets with $p_T > 30$ GeV and $|\eta_{\text{det}}| < 0.8$. The corresponding mistag rates, i.e. the probabilities to wrongly tag u,d,s,g jets, are 5% and 0.5%, respectively. This asymmetric tagging procedure was found to provide the best sensitivity to a Higgs boson signal. To cope with differences in track reconstruction efficiencies in data and in simulation, the b-tagging algorithm was applied directly only to the data, including the "multijet sample", while flavor-dependent tagging probabilities measured in dedicated data samples were applied to the simulated jets.

In the simulation, the (W/Z)+heavy-flavor jets backgrounds are normalized as follows. The next-to-leading order to leading order (NLO/LO) K-factors were determined with MCFM for (W/Z)+2-jets for both heavy and light flavors, and the ratio of those K-factors was applied to the (W/Z)+heavy-flavor jets events generated with ALPGEN, in addition to the overall normalization factor determined from data before b tagging. Figure 3 shows the leading di-jet invariant mass before and after b-tagging. The simulation gives a reasonable description of the data before b-tagging. After b-tagging there is some deficit at low di-jet masses. The numbers of signal and background events for type 1 and type 2 taus combined, before and after b-tagging requirements, are given in Table I. After b tagging; the W+jets and Z+jets backgrounds are dominated ($\mathcal{O}(90\%)$) by heavy flavor jet production.

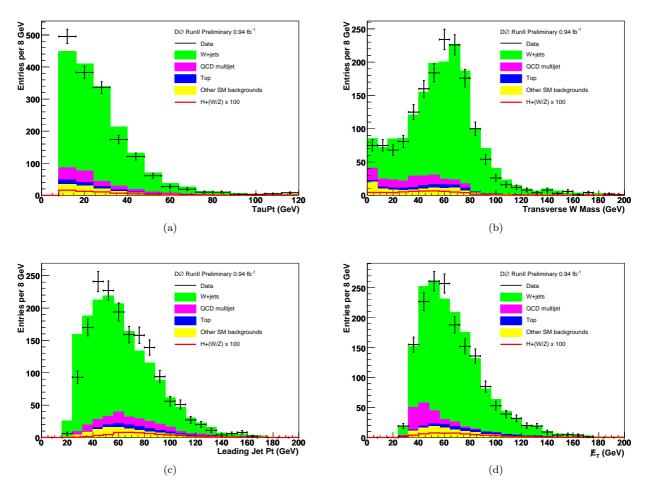


FIG. 2: Distributions for (a) $p_T(\tau)$ (b) m_T (c) $p_T(jet(1))$ (d) E_T , for tau type 1 and 2 events with 2 and 3 jets combined. These figures show the agreement between data and expected background in important variables for this analysis.

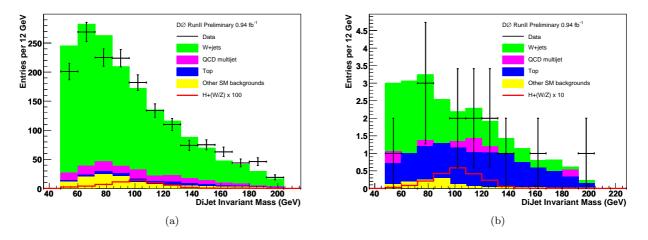


FIG. 3: Distributions in the analysis sample of the invariant mass of the two leading jets (a) before and (b) after b tagging for tau type 1 and 2 events with 2 and 3 jets combined. The data are shown as points with error bars. The various background contributions (SM and multijet) are shown as histograms, with color codes as indicated in the frames. Distributions for a signal (W/Z + H) with a Higgs boson mass of 115 GeV are also shown, multiplied by 100 before and 10 after b tagging.

TABLE II: Systematic uncertainties on the numbers of predicted signal and SM background events for uncertainties dependent on tau type, in percents: Tau reconstruction (τ ID), tau energy scale (τ ES), tau track efficiency (τ_{trk} ID), electron veto efficiency (ϵ veto) and multijet estimate (QCD). For the signal, a Higgs boson mass of 115 GeV is chosen.

	τ ID	$\tau \text{ ES}$	$ au_{trk}$ ID	e veto
Signal	5.0 - 10.0	1.7 - 2.3	1.0	_
SM backgrounds	5.0 - 10.0	1.7 - 2.3	1.0	$10.0 \ (\tau \ \text{type 2 only})$

TABLE III: Systematic uncertainties on the numbers of predicted signal and SM background events, in percents: integrated luminosity (Lumi), trigger, jet energy calibration (J-C), resolution (J-R), and reconstruction efficiency (J-E), b tagging, cross sections (σ), heavy flavor fractions (HF). For the signal, a Higgs boson mass of 115 GeV is chosen.

	Lumi	Trigger	J-C	J-R	J-E	b tagging	σ	HF
Signal	6.1	5.5	7.0	1.1	1.7	1.4	6	
SM backgrounds	6.1	5.5	1.0	4.0	4.9	4.4	6 - 18	30

IV. SYSTEMATIC UNCERTAINTIES

Experimental systematic uncertainties arise from the trigger parameterisation, from the jet energy calibration, jet resolution, jet reconstruction efficiency, tau energy calibration, track and reconstruction efficiency, and from b tagging. The multijet estimate also has an associated uncertainty which is derived for each tau type in the analysis and is large relative to other systematics due to the small statistics in the multijet estimate samples after b tagging. The uncertainty on the multijet background is evaluated to be between 35% - 100% for the three tau types. The parameterizations applied to the simulation were varied within their uncertainties, and the impact on both the normalizations and the shapes of the di-jet mass were assessed. The correlations between these systematic uncertainties in the signal and background processes were taken into account to derive the final results, as well as for the uncertainty on the integrated luminosity of the data sample.

The cross sections of the various SM processes have theoretical uncertainties. Uncertainties on the heavy flavor fractions can be partially obtained with MCFM. They were substantially inflated to cope with unaccounted effects, such as heavy quark masses. For the other background processes, theoretical uncertainties were taken from Ref. [13], or calculated with MCFM. The normalization of the multijet background is fully anticorrelated with the normalization of the SM backgrounds, as the sum is constrained by the data before b tagging. The uncertainties on the signal cross sections were taken from Ref. [14].

The systematic uncertainties affecting the signal and SM background normalizations that are dependent (independent) on tau type are reported in Table II(III).

V. RESULTS

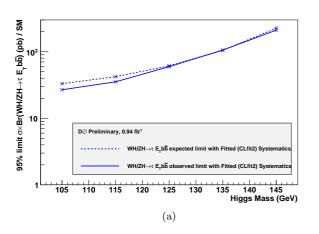
To set limits on the SM Higgs boson production cross section, a modified frequentist approach [17] was used. The signal confidence level CL_s , is defined as the ratio of the confidence level for the signal-plus-background hypothesis to the background-only hypothesis ($CL_s = CL_{s+b}/CL_b$). It is calculated by integration of the distributions of a test statistic over the outcomes of pseudo-experiments, generated according to Poisson statistics, for the signal+background and background-only hypotheses. The test statistic is calculated as a joint log-likelihood ratio (LLR) obtained by summing LLR values over the bins of the di-jet mass distribution. Systematic uncertainties were incorporated via Gaussian smearing of the Poisson probability distributions for signal and backgrounds within the pseudo-experiments. All correlations between signal and backgrounds were maintained. To reduce the impact of systematic uncertainties on the sensitivity of the analysis, the individual signal and background contributions were fitted to the data (and pseudo-data) for both the signal-plus-background and the background-only hypotheses independently by maximizing a profile likelihood function for each hypothesis [18]. The profile likelihood is constructed via a joint Poisson probability over the number of bins in the calculation, with nuisance parameters multiplying each of the components of systematic error in the calculation. The nuisance parameters are fit to signal plus background and background only hypotheses with Gaussian constraints on their prior predictions. The maximization of the likelihood function is performed by varying the nuisance parameters.

The ratio of the observed and expected cross section times branching ratio limits to the SM predictions is shown in Fig. 4 and in Table IV. The LLRs are also shown in Fig. 4. The observed limit is in agreement with the expected

TABLE IV: For various Higgs boson masses, observed and expected ratios of excluded to SM production cross sections times branching fraction for $H \to b\bar{b}$.

mass (GeV)	105	115	125	135	145
observed	26.8	35.4	59.6	106.0	211.2
expected	33.3	42.1	61.5	104.8	226.4

limit, defined as the $\pm 1\sigma$ median of the limits obtained in background-only pseudo experiments. For a 115 GeV Higgs boson mass, the observed and expected limits on the cross section of combined HW and HZ production times branching fraction for $H \to b\bar{b}$ are 35.4 and 42.1 times larger than the SM value, respectively.



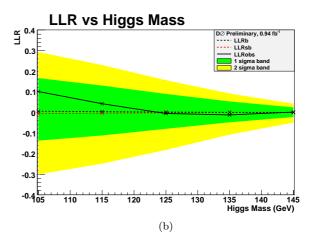


FIG. 4: As a function of the Higgs boson mass, (a) limit on the cross section of combined HW and HZ production times branching fraction for $H \to b\bar{b}$, relative to the standard model value, and (b) log likelihood ratio. In (a) the observed and expected limits are shown as solid and dashed curves, respectively. In (b) the observed LLR is shown as a solid curve, the expected LLRs are shown as black and red dashed curves for the background-only and signal+background hypotheses, respectively, and the green and yellow areas correspond to the one and two σ deviations around the expected background-only LLR.

VI. SUMMARY

A search for the standard model Higgs boson has been performed in 0.94 fb⁻¹ of $p\bar{p}$ collisions at 1.96 TeV. The topology analysed consists of a pair of b jets, a τ and E_T , as expected from $p\bar{p} \to HW \to b\bar{b}\tau\nu$. The search is also sensitive to HZ production, where the Z decays to taus and one tau is undetected. No deviation from the expectation from standard model backgrounds was observed. The di-jet mass of the event was used to derive an upper limit on the cross section of the $p\bar{p} \to HW$ and $p\bar{p} \to HZ$ processes combined, as a function of the Higgs boson mass. For a mass of 115 GeV, this limit is a factor 35 larger than the standard model cross section.

Acknowledgments

We thank the staffs at Fermilab and collaborating institutions, and acknowledge support from the DOE and NSF (USA); CEA and CNRS/IN2P3 (France); FASI, Rosatom and RFBR (Russia); CNPq, FAPERJ, FAPESP and FUNDUNESP (Brazil); DAE and DST (India); Colciencias (Colombia); CONACyT (Mexico); KRF and KOSEF (Korea); CONICET and UBACyT (Argentina); FOM (The Netherlands); STFC (United Kingdom); MSMT and GACR (Czech Republic); CRC Program, CFI, NSERC and WestGrid Project (Canada); BMBF and DFG (Germany); SFI (Ireland); The Swedish Research Council (Sweden); CAS and CNSF (China); and the Alexander von Humboldt

Foundation	(Germany).	

- [1] M. Carena et al., "Report of the Tevatron Higgs Working Group", arXiv:hep-ph/0010338; CDF and DØ Collaborations, "Results of the Tevatron Higgs Sensitivity Study", FERMILAB-PUB-03/320-E.
- [2] R. Barate et al. [LEP Working Group for Higgs boson searches], Phys. Lett. B 565, 61 (2003).
- [3] The LEP Collaborations: ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, the LEP Electroweak Working Group, "Precision Electroweak Measurements and Constraints on the Standard Model," CERN-PH-EP/2007-039, arXiv:0712.0929v2.
- [4] T. Scanlon, "b-Tagging and the Search for Neutral Supersymmetric Higgs Bosons at $D\emptyset$ ", FERMILAB-THESIS-2006-43.
- [5] V.M. Abazov et al. (D0 Collaboration), Nucl. Instrum. Methods in Phys. Res. A 565, 463 (2006).
- [6] T. Andeen et al., FERMILAB-TM-2365 (2007).
- [7] M.L. Mangano et al., JHEP **0307**, 001 (2003); version 2.05 (2.11) was used for Run IIa (Run IIb) simulation.
- [8] T. Sjöstrand, S. Mrenna and P. Skands, JHEP 0605, 026 (2006); version 6.323 (6.409) was used for Run IIa (Run IIb) simulation.
- [9] S. Höche et al., "Matching Parton Showers and Matrix Elements", in Proceedings of the Workshop on the Implications of HERA for LHC Physics, ed. A. De Roeck and H. Jung (CERN, Geneva, 2005), p288.
- [10] E. Boos et al. (CompHEP Collaboration), Nucl. Instrum. Methods in Phys. Res. A 534, 250 (2004).
- [11] J. Pumplin et al., JHEP **0207**, 012 (2002); D. Stump et al., JHEP **0310**, 046 (2003).
- [12] J.M. Campbell and R.K. Ellis, Phys. Rev. D 60, 113006 (1999).
- [13] M. Cacciari et al., JHEP 404, 068 (2004); N. Kidonakis and R. Vogt, Phys. Rev. D 68, 114014 (2003); N. Kidonakis, Phys. Rev. D 74, 114012 (2006).
- [14] S. Catani et al., JHEP 0307, 028 (2003).
- [15] R. Brun and F. Carminati, CERN Program Library Long Writeup W5013, 1993 (unpublished).
- [16] The DØ Collaboration, "Search for WH Production using a Neural Network Approach in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV with 1.7 fb⁻¹ of Data", DØ Note 5472-CONF.
- [17] T. Junk, Nucl. Instrum. Methods in Phys. Res. A 434, 435 (1999); A. Read, in "1st Workshop on Confidence Limits," CERN Report No. CERN-2000-005, 2000.
- [18] W. Fisher, FERMILAB-TM-2386-E.